Partial English Translation of Japanese Laid-Open Patent Application No. 9-265482

[0111]

Embodiment 14, Figs. 55 to 59 are diagrams explaining the database contents information acquisition means according to still another embodiment of the present invention. Fig. 55 is a diagram illustrating a constitution of the device according to this embodiment, Fig. 56 is a flowchart illustrating a flow of an operation of the acquisition abnormality determination means 11 in Fig. 55, Fig. 57 is a diagram illustrating an example of concrete data of a response header to HTTP request, Fig. 58 is a flowchart illustrating a flow of an operation of the abnormality information deletion means 12, and Fig. 59 is a diagram illustrating an example of data stored in the acquisition target address-related information storage means 42d in Fig. 55 for storing a number of times at which acquisition is determined as abnormal. The constitution in this embodiment is provided with an acquisition abnormality determination means 61 for determining acquisition abnormality, an abnormality information deletion means 62 for deleting information determined as abnormal from acquired information or acquisition target address-related information, and an abnormality information deletion condition input means 63 for inputting a condition for deleting abnormality information as shown in Fig. 55 in addition to the constitution of the embodiment 10 shown in Fig. 44. Further, additional information is also stored in

1

an acquisition target address-related information storage means 42d. The acquisition abnormality determination means 61 and the abnormality information deletion means 62 are generally realized by software which operates on a computer system, and the abnormality information deletion condition input means 13 is realized by units such as a mouse, a keyboard and a display.

[0112]

The operation is explained below with reference to the drawings, but a difference in the operation between this embodiment and the embodiment 10 is an acquisition target address-related information storage means 42d, an acquisition abnormality determination means 61, an abnormality information deletion means 62, and an abnormality information deletion condition input means 63. The operation in these means is mainly explained. In this operation, these means discriminate meaningless information acquired at the time of communication error frequently occurring due to malfunction of network or the like from the acquired information, do not acquire the meaningless information temporarily so as to prevent the unnecessary information from being stored as a retrieval target in the database. In the case where the communication error successively occurs, these means remove the URL from an acquisition target, so as to prevent acquisition thereafter.

[0113]

In HTTP communication, when information is acquired, information relating to a communication state can be acquired in a format of short data which is called as a response header.

2

Three examples of the response header are shown in Figs. 57(a) to 57(c). Numbers written on the first lines of the respective response headers are code number of the response headers, and these numbers indicate success/failure of communication. 200s indicate success of communication, 300s indicate that an address for requesting communication transfers and its transfer destination, 400s indicate that the way communication is requested is wrong, and 500s indicate that the server cannot process the communication request. Like the example of Fig. 57(a), therefore, when the code number is 200s, the acquired result can be regarded as normal, and like the examples of Figs. 57(b) and 57(c), when the code number is 400s or 500s, the acquired result can be regarded as abnormal. Not shown in the examples, 300s may be regarded as normal.

[0114]

After the meaning of the response headers is taken into consideration, the acquired information is determined as normal or abnormal. One example of such determination is shown in a processing flowchart of the acquisition abnormality determining means 61 shown in Fig. 56. At step S131, the code number is read form the HTTP response header in the acquired information acquired by the acquisition means 44. At step S132, a check is made whether the code number is 200s or 300s. For example, in the example of Fig. 57, in the case of Figs. 57(b) and 57(c), since the code number is not 200s nor 300s, the sequence goes to step S133, so that the acquired result is determined as abnormal. Further, in the case of Fig. 57(a),

since the code number is 200s, the sequence goes to step S134, so that the acquired result is determined as normal. The method shown in Fig. 56 is one example of the abnormality determining method, and a method of executing processes other than that in Fig. 56, such as the processes of 400s and 500s, separately can be considered.

[0115]

One example of the operation of the abnormality information deletion means 62 is explained below with reference to a flowchart of Fig. 58, but the abnormality information deletion means 62 mainly functions when the acquisition is repeated. At step S151, an address where the acquired information is determined as abnormal by the acquisition abnormality determination unit 61 is read. At step S152, a numerical value which indicates a number of occurrence times of acquisition abnormality relating to URL in the acquisition target address-related information storage means 42d is increased. Fig. 59 illustrates an example of the data to be stored in the acquisition target address-related information storage means 42d. At step S153, the abnormality information is deleted from the database contents information storage means 4. At step S154, a standard number of deletion times of abnormality information input by the abnormality information deletion condition input means 63 is read. In the case where determination of abnormality continues at a certain number of times, when the abnormality is not caused simply by temporary convergence of a line or temporary shut-down of a server, this

standard number of deletion times is set in order to determine that the abnormality is caused by unpresent of an address whose information is requested or that communication with the server is unstable. At S155, the number of abnormality times stored in the acquisition target address-related information storage means 42d is compared with the basic number of deletion times of the abnormality information as to whether the data of the address determined as abnormality are continuously determined as abnormal at more than the standard number of times. When the number of abnormality determination times exceeds the standard value, the sequence goes to step S156, and the address is removed form the acquisition target. That is to say, when the address is present in the acquisition target address-related information storage means 42d, the address is deleted, but when it is not present, it is not registered. When the continuation of abnormality does not exceed the standard value, the sequence goes to step S157, so that the address is left in the acquisition target address-related information storage means 42d.

[0116]

According to the embodiment, abnormality information in the acquired information can be determined, and unnecessary information can be deleted suitably. In the case where the acquisition abnormality continues, the address can be automatically removed from the acquisition target. As a result, only necessary information can be acquired efficiently.

FIG. 55

41: ACQUISITION START CONDITION INPUT MEANS

42D: ACQUISITION TARGET ADDRESS-RELATED INFORMATION STORAGE

MEANS

43: ACQUISITION DESTINATION DETERMINATION MEANS

44: ACQUISITION MEANS

61: ACQUISITION ABNORMALITY DETERMINATION MEANS

62: ABNORMALITY INFORMATION DELETION MEANS

63: ABNORMALITY INFORMATION DELETION CONDITION INPUT MEANS

2: COMMUNICATION CONTROL DEVICE

4: DATABASE CONTENTS INFORMATION STORAGE MEANS

1: EXTERNAL DATABASE

47: LINK DESTINATION ADDRESS EXTRACTION MEANS

48: ACQUISITION TARGET LIMIT CONDITION INPUT MEANS

49: ACQUISITION TARGET LIMIT MEANS

FIG. 56

S131: READ HEADER CODE NUMBER OF HTTP RESPONSE OF ACQUISITION

INFORMATION

S132: IS THE HEADER CODE NUMBER IS 200S OR 300S?

S133: DETERMINE ACQUISITION RESULT AS ABNORMALITY

CODE IS 200S OR 300S

CODE IS NOT 200S OR 300S

DETERMINE ACQUISITION RESULT AS NORMAL

FIG. 57

(a)

HTTP/1.0 200 Document follows

MIME-Version: 1.0

6

Server: CERN/3.0

Date: Wednesday, 18-Sep-96 05:40:51 GMT

Content-Type: text/html

Last-Modified: Friday, 30-Aug-96 07:03-05 GMT

(b)

HTTP/1.0 404 No Such file or directory.

Content-Type: text/html

Server:Spinner/1.0b12

Date: Wed, 25 Sep 1996 07:18:05+0000

Content-Length: 634

(c)

HTTP/1.0500 Error from proxy

Proxy-agent: Netscape-Proxy/1.12

Content-Type: text/html

FIG. 58

S151: READ ADDRESS WHERE ACQUIRED INFORMATION IS DETERMINED AS ABNORMAL

S152: INCREASE NUMBER OF ABNORMALITY COUNTER OF ACQUISITION ADDRESS-RELATED INFORMATION STORAGE MEANS RELATING TO THE ADDRESS

S153: DELETE ACQUIRED INFORMATION OF ADDRESS

S154: READ STANDARD NUMBER OF TIMES OF ABNORMALITY DETERMINATION

S155: DOES CONTINUATION OF ABNORMALITY EXCEED STANDARD VALUE?

S156: REMOVE FROM ACCUSATION TARGET

CONTINUATION OF ABNORMALITY DETERMINATION IS LESS THAN STANDARD NUMBER OF TIMES

ABNORMALITY DETERMINATION CONTINUES AT MORE THAN STANDARD
NUMBER OF TIMES

S157: ADDRESS IS SET AS ACQUISITION TARGET

FIG. 59

MANAGEMENT NUMBER

(UNIT ID-ID IN UNIT)

ACQUISITION TARGET ADDRESS

(URL ABSOLUTE ADDRESS)

ACQUISITION CHECK

DONE

UNDONE

NUMBER OF ABNORMALITY NUMBERS